Markov Chains

Gonzalo Mateos
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.ece.rochester.edu/~gmateosb/

October 3, 2018

Limiting distributions

Ergodicity

Queues in communication networks: Limit probabilities

Limiting distributions

- MCs have one-step memory. Eventually they forget initial state
- Q: What can we say about probabilities for large n?

$$\pi_j := \lim_{n\to\infty} \mathrm{P}\left(X_n = j \,\middle|\, X_0 = i\right) = \lim_{n\to\infty} P_{ij}^n$$

$\Rightarrow$ Assumed that limit is independent of initial state $X_0 = i$

- We've seen that this problem is related to the matrix power $\mathbf{P}^n$

$$\mathbf{P} = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}, \qquad \mathbf{P}^7 = \begin{pmatrix} 0.6031 & 0.3969 \\ 0.5953 & 0.4047 \end{pmatrix}$$

$$\mathbf{P}^2 = \begin{pmatrix} 0.7 & 0.3 \\ 0.45 & 0.55 \end{pmatrix}, \qquad \mathbf{P}^{30} = \begin{pmatrix} 0.6000 & 0.4000 \\ 0.6000 & 0.4000 \end{pmatrix}$$

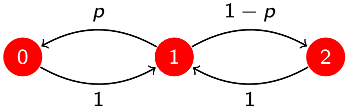
- Matrix product converges $\Rightarrow$ probs. independent of time (large n)
- All rows are equal $\Rightarrow$ probs. independent of initial condition

Periodicity

- **Def:** Period d of a state i is (gcd means greatest common divisor)

$$d = \gcd\left\{n : P_{ii}^n \neq 0\right\}$$

- State i is periodic with period d if and only if
  - $\Rightarrow P_{ii}^n \neq 0$ only if n is a multiple of d
  - $\Rightarrow d$ is the largest number with this property
- Positive probability of returning to i only every d time steps
  - $\Rightarrow$ If period $d = 1$ state is aperiodic (most often the case)
  - $\Rightarrow$ Periodicity is a class property



- State 1 has period 2. So do 0 and 2 (class property)
- Ex: One dimensional random walk also has period 2

Periodicity example

Example

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} 0.50 & 0.50 \\ 0.25 & 0.75 \end{pmatrix}, \quad \mathbf{P}^3 = \begin{pmatrix} 0.250 & 0.750 \\ 0.375 & 0.625 \end{pmatrix}$$

- $P_{11} = 0$, but $P_{11}^2, P_{11}^3 \neq 0$ so $\gcd\{2, 3, \ldots\} = 1$. State 1 is aperiodic
- $P_{22} \neq 0$. State 2 is aperiodic (had to be, since $1 \leftrightarrow 2$)

Example

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{P}^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ldots$$

- $P_{11}^{2n+1} = 0$, but $P_{11}^{2n} \neq 0$ so $\gcd\{2, 4, \ldots\} = 2$. State 1 has period 2
- The same is true for state 2 (since $1 \leftrightarrow 2$)

Positive recurrence and ergodicity

- **Recall:** state i is recurrent if the MC returns to i with probability 1
  - $\Rightarrow$ Define the return time to state i as

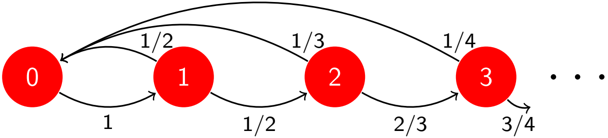
$$T_i = \min\{n > 0 : X_n = i \,\big|\, X_0 = i\}$$

- **Def:** State i is positive recurrent when expected value of T_i is finite

$$\mathbb{E}\left[T_i \,\big|\, X_0 = i\right] = \sum_{n=1}^{\infty} n\mathrm{P}\left(T_i = n \,\big|\, X_0 = i\right) < \infty$$

- **Def:** State i is null recurrent if recurrent but $\mathbb{E}\left[T_i \,\big|\, X_0 = i\right] = \infty$
  - $\Rightarrow$ Positive and null recurrence are class properties
  - $\Rightarrow$ Recurrent states in a finite-state MC are positive recurrent
- **Def:** Jointly positive recurrent and aperiodic states are ergodic
  - $\Rightarrow$ Irreducible MC with ergodic states is said to be an ergodic MC

Null recurrent Markov chain example



$$P\left(T_0 = 2 \,\middle|\, X_0 = 0\right) = \frac{1}{2} \qquad P\left(T_0 = 3 \,\middle|\, X_0 = 0\right) = \frac{1}{2} \times \frac{1}{3}$$

$$P\left(T_0 = 4 \,\middle|\, X_0 = 0\right) = \frac{1}{2} \times \frac{2}{3} \times \frac{1}{4} = \frac{1}{3 \times 4} \quad \ldots \quad P\left(T_0 = n \,\middle|\, X_0 = 0\right) = \frac{1}{(n-1) \times n}$$

- State 0 is recurrent because probability of not returning is 0

$$P\left(T_0 = \infty \,\middle|\, X_0 = 0\right) = \lim_{n \to \infty} \frac{1}{(n-1) \times n} \to 0$$

- Also null recurrent because expected return time is infinite

$$\mathbb{E}\left[T_0 \,\middle|\, X_0 = 0\right] = \sum_{n=2}^{\infty} n P\left(T_0 = n \,\middle|\, X_0 = 0\right) = \sum_{n=2}^{\infty} \frac{1}{(n-1)} = \infty$$

Limit distribution of ergodic Markov chains

Theorem

For an ergodic (i.e., irreducible, aperiodic and positive recurrent) MC, $\lim_{n\to\infty} P_{ij}^n$ *exists and is independent of the initial state* i*, i.e.,*

$$\pi_j = \lim_{n\to\infty} P_{ij}^n$$

Furthermore, steady-state probabilities $\pi_j \geq 0$ *are the unique nonnegative solution of the system of linear equations*

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \qquad \sum_{j=0}^{\infty} \pi_j = 1$$

- Limit probs. independent of initial condition exist for ergodic MC
  - $\Rightarrow$ Simple algebraic equations can be solved to find π_j
- No periodic, transient, null recurrent states, or multiple classes

Algebraic relation to determine limit probabilities

- Difficult part of theorem is to prove that $\pi_j = \lim_{n\to\infty} P_{ij}^n$ exists
- To see that algebraic relation is true use total probability

$$
\begin{aligned}
P_{kj}^{n+1} &= \sum_{i=0}^{\infty} \mathrm{P}\left(X_{n+1} = j \,\middle|\, X_n = i, X_0 = k\right) P_{ki}^n \\
&= \sum_{i=0}^{\infty} P_{ij} P_{ki}^n
\end{aligned}
$$

- If limits exists, $P_{kj}^{n+1} \approx \pi_j$ and $P_{ki}^n \approx \pi_i$ (sufficiently large n)

$$
\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}
$$

- The other equation is true because the π_j are probabilities

Vector/matrix notation: Matrix limit

- More compact and illuminating using vector/matrix notation
  - $\Rightarrow$ Finite MC with J states
- First part of theorem says that $\lim_{n\to\infty} \mathbf{P}^n$ exists and

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_J \\ \pi_1 & \pi_2 & \dots & \pi_J \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_J \end{pmatrix}$$

- Same probabilities for all rows $\Rightarrow$ Independent of initial state
- Probability distribution for large n

$$\lim_{n\to\infty} \mathbf{p}(n) = \lim_{n\to\infty} (\mathbf{P}^T)^n \mathbf{p}(0) = [\pi_1, \dots, \pi_J]^T$$

  - $\Rightarrow$ Independent of initial condition $\mathbf{p}(0)$

Vector/matrix notation: Eigenvector

- **Def:** Vector limit (steady-state) distribution is $\boldsymbol{\pi} := [\pi_1, \ldots, \pi_J]^T$
- Limit distribution is unique solution of ($\mathbf{1} := [1, 1, \ldots]^T$)

$$\boldsymbol{\pi} = \mathbf{P}^T \boldsymbol{\pi}, \qquad \boldsymbol{\pi}^T \mathbf{1} = 1$$

- $\boldsymbol{\pi}$ eigenvector associated with eigenvalue 1 of $\mathbf{P}^T$
  - Eigenvectors are defined up to a scaling factor
  - Normalize to sum 1
- All other eigenvalues of $\mathbf{P}^T$ have modulus smaller than 1
  - If not, $\mathbf{P}^n$ diverges, but we know $\mathbf{P}^n$ contains n-step transition probs.
  - $\boldsymbol{\pi}$ eigenvector associated with largest eigenvalue of $\mathbf{P}^T$
- Computing $\boldsymbol{\pi}$ as eigenvector is often computationally efficient

Vector/matrix notation: Rank

- Can also write as ($\mathbf{I}$ is identity matrix, $\mathbf{0} = [0, 0, \ldots]^T$)

$$\left(\mathbf{I} - \mathbf{P}^T\right) \boldsymbol{\pi} = \mathbf{0} \qquad \boldsymbol{\pi}^T \mathbf{1} = 1$$

- $\boldsymbol{\pi}$ has J elements, but there are $J+1$ equations $\Rightarrow$ Overdetermined
- If 1 is eigenvalue of $\mathbf{P}^T$, then 0 is eigenvalue of $\mathbf{I} - \mathbf{P}^T$
  - $\mathbf{I} - \mathbf{P}^T$ is rank deficient, in fact $\text{rank}(\mathbf{I} - \mathbf{P}^T) = J - 1$
  - Then, there are in fact only J linearly independent equations
- $\boldsymbol{\pi}$ is eigenvector associated with eigenvalue 0 of $\mathbf{I} - \mathbf{P}^T$
  - $\boldsymbol{\pi}$ spans null space of $\mathbf{I} - \mathbf{P}^T$ (not much significance)

Ergodic Markov chain example

- MC with transition probability matrix

$$\mathbf{P} = \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0.1 & 0.5 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

- Q: Does $\mathbf{P}$ correspond to an ergodic MC?
  - Irreducible: all states communicate with state 2 $\checkmark$
  - Positive recurrent: irreducible and finite $\checkmark$
  - Aperiodic: period of state 2 is 1 $\checkmark$
- Then, there exist π_1, π_2 and π_3 such that $\pi_j = \lim_{n\to\infty} P_{ij}^n$

  $\Rightarrow$ Limit is independent of i

Ergodic Markov chain example (continued)

- Q: How do we determine the limit probabilities π_j?
- Solve system of linear equations $\pi_j = \sum_{i=1}^{3} \pi_i P_{ij}$ and $\sum_{j=1}^{3} \pi_j = 1$

$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0.1 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.7 & 0.4 & 0.7 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}$$

$\Rightarrow$ The blue block in the matrix above is $\mathbf{P}^T$

- There are three variables and four equations
  - Some equations might be linearly dependent
  - Indeed, summing first three equations: $\pi_1 + \pi_2 + \pi_3 = \pi_1 + \pi_2 + \pi_3$
  - Always true, because probabilities in rows of $\mathbf{P}$ sum up to 1
  - A manifestation of the rank deficiency of $\mathbf{I} - \mathbf{P}^T$
- Solution yields $\pi_1 = 0.0909$, $\pi_2 = 0.2987$ and $\pi_3 = 0.6104$

Stationary distribution

- Limit distributions are sometimes called stationary distributions
  - $\Rightarrow$ Select initial distribution to $\mathrm{P}(X_0 = i) = \pi_i$ for all i
- Probabilities at time $n = 1$ follow from law of total probability

$$\mathrm{P}(X_1 = j) = \sum_{i=1}^{\infty} \mathrm{P}\left(X_1 = j \,\middle|\, X_0 = i\right) \mathrm{P}(X_0 = i)$$

- Definitions of P_{ij}, and $\mathrm{P}(X_0 = i) = \pi_i$. Algebraic property of π_j

$$\mathrm{P}(X_1 = j) = \sum_{i=1}^{\infty} P_{ij}\pi_i = \pi_j$$

  - $\Rightarrow$ Probability distribution is unchanged
- Proceeding recursively, system initialized with $\mathrm{P}(X_0 = i) = \pi_i$
  - $\Rightarrow$ Probability distribution invariant: $\mathrm{P}(X_n = i) = \pi_i$ for all n
- MC stationary in a probabilistic sense (states change, probs. do not)

UNIVERSITY of ROCHESTER

Limiting distributions

Ergodicity

Queues in communication networks: Limit probabilities

Ergodicity

- **Def:** Fraction of time $T_i^{(n)}$ spent in i-th state by time n is

$$T_i^{(n)} := \frac{1}{n}\sum_{m=1}^{n} \mathbb{I}\left\{X_m = i\right\}$$

- Compute expected value of $T_i^{(n)}$

$$\mathbb{E}\left[T_i^{(n)}\right] = \frac{1}{n}\sum_{m=1}^{n} \mathbb{E}\left[\mathbb{I}\left\{X_m = i\right\}\right] = \frac{1}{n}\sum_{m=1}^{n} \mathrm{P}\left(X_m = i\right)$$

- As $n \to \infty$, probabilities $\mathrm{P}\left(X_m = i\right) \to \pi_i$ (ergodic MC). Then

$$\lim_{n\to\infty} \mathbb{E}\left[T_i^{(n)}\right] = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} \mathrm{P}\left(X_m = i\right) = \pi_i$$

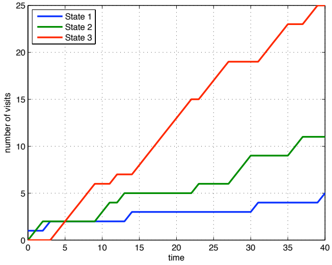
- For ergodic MCs same is true without expected value $\Rightarrow$ Ergodicity

$$\lim_{n\to\infty} T_i^{(n)} = \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} \mathbb{I}\left\{X_m = i\right\} = \pi_i, \quad \text{a.s.}$$

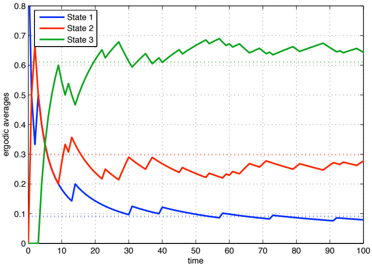
- Recall transition probability matrix

$$\mathbf{P} := \begin{pmatrix} 0 & 0.3 & 0.7 \\ 0.1 & 0.5 & 0.4 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

Visits to states, $nT_i^{(n)}$



Ergodic averages, $T_i^{(n)}$



- Ergodic averages slowly converge to $\boldsymbol{\pi} = [0.09, 0.29, 0.61]^T$

Function's ergodic average

Theorem

Consider an ergodic Markov chain with states $X_n = 0, 1, 2, \ldots$ and stationary probabilities π_j. Let $f(X_n)$ be a bounded function of the state X_n. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) = \sum_{j=1}^{\infty} f(j)\pi_j, \quad a.s.$$

- Ergodic average $\to$ Expectation under stationary distribution $\boldsymbol{\pi}$
- Use of ergodic averages is more general than $T_i^{(n)}$

  $\Rightarrow$ $T_i^{(n)}$ is a particular case with $f(X_m) = \mathbb{I}\{X_m = i\}$
- Think of $f(X_m)$ as a reward (or cost) associated with state X_m

  $\Rightarrow$ $(1/n)\sum_{m=1}^{n} f(X_m)$ is the time average of rewards (costs)

Function's ergodic average (cheat's proof)

Proof.

- Because $\mathbb{I}\{X_m = i\} = 1$ if and only if $X_m = i$ we can write

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) = \frac{1}{n}\sum_{m=1}^{n}\left(\sum_{i=1}^{\infty} f(i)\mathbb{I}\{X_m = i\}\right)$$

- Change order of summations. Use definition of $T_i^{(n)}$

$$\frac{1}{n}\sum_{m=1}^{n} f(X_m) = \sum_{i=1}^{\infty} f(i)\left(\frac{1}{n}\sum_{m=1}^{n}\mathbb{I}\{X_m = i\}\right) = \sum_{i=1}^{\infty} f(i)\,T_i^{(n)}$$

- Let $n \to \infty$, use ergodicity result for $\lim_{n\to\infty} T_i^{(n)} = \pi_i$ [cf. page 17]

$\square$

Ensemble and ergodic averages

- Ensemble average: across different realizations of the MC

$$\mathbb{E}\left[f(X_n)\right] = \sum_{i=1}^{\infty} f(i)\mathrm{P}\left(X_n = i\right) \to \sum_{i=1}^{\infty} f(i)\pi_i$$

- Ergodic average: across time for a single realization of the MC

$$\bar{f}_n = \frac{1}{n}\sum_{m=1}^{n} f(X_m)$$

- These quantities are fundamentally different
  $\Rightarrow$ But $\mathbb{E}\left[f(X_n)\right] = \bar{f}_n$ almost surely, asymptotically in n

- One realization of the MC as informative as all realizations
  $\Rightarrow$ Practical value: observe/simulate only one path of the MC

Ergodicity in periodic Markov chains

- Ergodic averages still converge if the MC is periodic
- For irreducible, positive recurrent MC (periodic or aperiodic) define

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \qquad \sum_{j=0}^{\infty} \pi_j = 1$$

- **Claim 1:** A unique solution exists (we say π_j are well defined)
- **Claim 2:** The fraction of time spent in state i converges to π_i

$$\lim_{n\to\infty} T_i^{(n)} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}\left\{X_m = i\right\} \to \pi_i$$

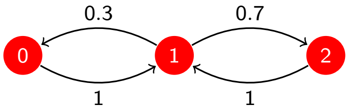
- If MC is periodic the probabilities P_{ij}^n oscillate

  $\Rightarrow$ But fraction of time spent in state i converges to π_i

Periodic irreducible Markov chain example

- Matrix **P** and state transition diagram of a periodic MC

$$\mathbf{P} := \begin{pmatrix} 0 & 1 & 0 \\ 0.3 & 0 & 0.7 \\ 0 & 1 & 0 \end{pmatrix}$$



- MC has period 2. If initialized with $X_0 = 1$, then

$$\begin{aligned} P_{11}^{2n+1} &= \mathrm{P}\left(X_{2n+1} = 1 \,\middle|\, X_0 = 1\right) = 0, \\ P_{11}^{2n} &= \mathrm{P}\left(X_{2n} = 1 \,\middle|\, X_0 = 1\right) = 1 \neq 0 \end{aligned}$$

- Define $\boldsymbol{\pi} := [\pi_1, \pi_2, \pi_3]^T$ as solution of

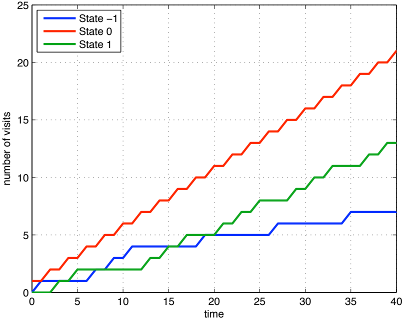
$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0.3 & 0 \\ 1 & 0 & 1 \\ 0 & 0.7 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}$$

$\Rightarrow$ Normalized eigenvector for eigenvalue 1 ($\boldsymbol{\pi} = \mathbf{P}^T\boldsymbol{\pi}$, $\boldsymbol{\pi}^T\mathbf{1} = 1$)

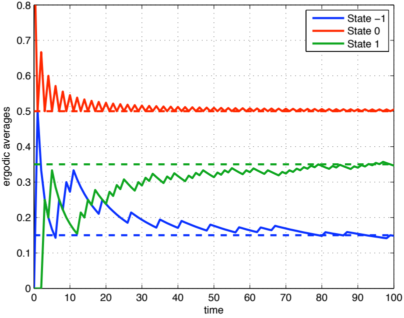
Periodic irreducible MC example (continued)

- Solution yields $\pi_1 = 0.15$, $\pi_2 = 0.50$ and $\pi_3 = 0.35$

Visits to states, $nT_i^{(n)}$



Ergodic averages, $T_i^{(n)}$



- Ergodic averages $T_i^{(n)}$ converge to the ergodic limits π_i

Periodic irreducible MC example (continued)

- Powers of the transition probability matrix do not converge

$$\mathbf{P}^2 = \begin{pmatrix} 0.3 & 0 & 0.7 \\ 0 & 1 & 0 \\ 0.3 & 0 & 0.7 \end{pmatrix}, \qquad \mathbf{P}^3 = \begin{pmatrix} 0 & 1 & 0 \\ 0.3 & 0 & 0.7 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{P}$$

$\Rightarrow$ In general we have $\mathbf{P}^{2n} = \mathbf{P}^2$ and $\mathbf{P}^{2n+1} = \mathbf{P}$

- At least one other eigenvalue of $\mathbf{P}^T$ has modulus 1

$$\left|\text{eig}_2\left(\mathbf{P}^T\right)\right| = 1$$

$\Rightarrow$ In this example, eigenvalues of $\mathbf{P}^T$ are 1, -1 and 0

Reducible Markov chains

- If MC is not irreducible it can be decomposed in transient ($\mathcal{T}_k$), ergodic ($\mathcal{E}_k$), periodic ($\mathcal{P}_k$) and null recurrent ($\mathcal{N}_k$) components
  - $\Rightarrow$ All these are (communication) class properties
- Limit probabilities for transient states are null

$$\mathrm{P}\left(X_n = i\right) \to 0, \text{ for all } i \in \mathcal{T}_k$$

- For arbitrary ergodic component $\mathcal{E}_k$, define conditional limits

$$\pi_j = \lim_{n\to\infty} \mathrm{P}\left(X_n = j \,\middle|\, X_0 \in \mathcal{E}_k\right), \quad \text{for all } j \in \mathcal{E}_k$$

- Results in pages 8 and 19 are true with this (re)defined π_j, where

$$\pi_j = \sum_{i\in\mathcal{E}_k} \pi_i P_{ij}, \quad \sum_{j\in\mathcal{E}_k} \pi_j = 1, \quad \text{for all } j \in \mathcal{E}_k$$

Reducible Markov chains (continued)

- Likewise, for arbitrary periodic component $\mathcal{P}_k$ (re)define π_j as

$$\pi_j = \sum_{i \in \mathcal{P}_k} \pi_i P_{ij}, \quad \sum_{j \in \mathcal{P}_k} \pi_j = 1, \quad \text{for all } j \in \mathcal{P}_k$$

- Probabilities $\mathrm{P}\left(X_n = j \,\middle|\, X_0 \in \mathcal{P}_k\right)$ do not converge (they oscillate)
- A conditional version of the result in page 22 is true

$$\lim_{n \to \infty} T_i^{(n)} := \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{I}\left\{X_m = i \,\middle|\, X_0 \in \mathcal{P}_k\right\} \to \pi_i$$

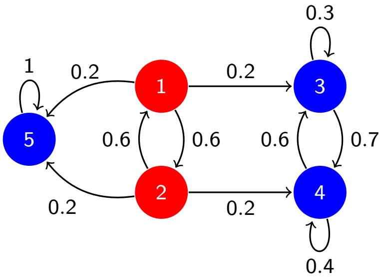
- Limit probabilities for null-recurrent states are null

$$\mathrm{P}\left(X_n = i\right) \to 0, \;\text{ for all } i \in \mathcal{N}_k$$

Reducible Markov chain example

- Transition matrix and state diagram of a reducible MC

$$\mathbf{P} := \begin{pmatrix} 0 & 0.6 & 0.2 & 0 & 0.2 \\ 0.6 & 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$



- States 1 and 2 are transient $\mathcal{T} = \{1, 2\}$
- States 3 and 4 form an ergodic class $\mathcal{E}_1 = \{3, 4\}$
- State 5 (absorbing) is a separate ergodic class $\mathcal{E}_2 = \{5\}$

Reducible MC example - Matrix powers

- 5-step and 10-step transition probabilities

$$\mathbf{P}^5 = \begin{pmatrix} 0 & 0.08 & 0.24 & 0.22 & 0.46 \\ 0.08 & 0 & 0.19 & 0.27 & 0.46 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{P}^{10} = \begin{pmatrix} 0.00 & 0 & 0.23 & 0.27 & 0.50 \\ 0 & 0.00 & 0.23 & 0.27 & 0.50 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0.46 & 0.54 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Transition into transient states is vanishing (columns 1 and 2)

  $\Rightarrow$ From $\mathcal{T} = \{1, 2\}$ will end up in either $\mathcal{E}_1 = \{3, 4\}$ or $\mathcal{E}_2 = \{5\}$

- Transition from 3 and 4 into 3 and 4 only

  $\Rightarrow$ If initialized in ergodic class $\mathcal{E}_1 = \{3, 4\}$ stays in $\mathcal{E}_1$

- Transition from 5 only into 5 (absorbing state)

Reducible MC example - Matrix decomposition

- Matrix $\mathbf{P}$ can be decomposed in blocks

$$\mathbf{P} = \begin{pmatrix} 0 & 0.6 & 0.2 & 0 & 0.2 \\ 0.6 & 0 & 0 & 0.2 & 0.2 \\ 0 & 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 0.6 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{\mathcal{T}} & \mathbf{P}_{\mathcal{T}\mathcal{E}_1} & \mathbf{P}_{\mathcal{T}\mathcal{E}_2} \\ \mathbf{0} & \mathbf{P}_{\mathcal{E}_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{\mathcal{E}_2} \end{pmatrix}$$

  (a) Block $\mathbf{P}_{\mathcal{T}}$ describes transition between transient states
  (b) Blocks $\mathbf{P}_{\mathcal{E}_1}$ and $\mathbf{P}_{\mathcal{E}_2}$ describe transitions within ergodic components
  (c) Blocks $\mathbf{P}_{\mathcal{T}\mathcal{E}_1}$ and $\mathbf{P}_{\mathcal{T}\mathcal{E}_2}$ describe transitions from $\mathcal{T}$ to $\mathcal{E}_1$ and $\mathcal{E}_2$

- Powers of n can be written as

$$\mathbf{P}^n = \begin{pmatrix} \mathbf{P}_{\mathcal{T}}^n & \mathbf{Q}_{\mathcal{T}\mathcal{E}_1} & \mathbf{Q}_{\mathcal{T}\mathcal{E}_2} \\ \mathbf{0} & \mathbf{P}_{\mathcal{E}_1}^n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}_{\mathcal{E}_2}^n \end{pmatrix}$$

- The transient transition block vanishes, $\lim_{n\to\infty} \mathbf{P}_{\mathcal{T}}^n = \mathbf{0}$

Reducible MC example - Limiting behavior

- As n grows the MC hits an ergodic state almost surely
  - $\Rightarrow$ Henceforth, MC stays within ergodic component

$$\mathrm{P}\left(X_{n+m} \in \mathcal{E}_i \,\middle|\, X_n \in \mathcal{E}_i\right) = 1, \quad \text{for all } m$$

- For large n suffices to study ergodic components
  - $\Rightarrow$ Behaves like a MC with transition probabilities $\mathbf{P}_{\mathcal{E}_1}$
  - $\Rightarrow$ Or like one with transition probabilities $\mathbf{P}_{\mathcal{E}_2}$
- We can think of all MCs as ergodic
- Ergodic behavior cannot be inferred a priori (before observing)
- Becomes known a posteriori (after observing sufficiently large time)

Cultural aside: Something is known a priori if its knowledge is independent of experience (MCs exhibit ergodic behavior). A posteriori knowledge depends on experience (values of the ergodic limits). They are inherently different forms of knowledge (search for Immanuel Kant).

Limiting distributions

Ergodicity

Queues in communication networks: Limit probabilities

Non-concurrent communication queue

- Communication system: Move packets from source to destination
- Between arrival and transmission hold packets in a memory buffer
- Example engineering problem, buffer design:
  - Packets arrive at a rate of 0.45 packets per unit of time
  - Packets depart at a rate of 0.55 packets per unit of time
  - How big should the buffer be to have a drop rate smaller than 10^{-6}? (i.e., one packet dropped for every million packets handled)
- Model: Time slotted in intervals of duration Δt. Each time slot n

  $\Rightarrow$ A packet arrives with prob. λ, arrival rate is $\lambda/\Delta t$

  $\Rightarrow$ A packet is transmitted with prob. μ, departure rate is $\mu/\Delta t$
- No concurrence: No simultaneous arrival and departure (small Δt)

Queue evolution equations (reminder)

- Q_n denotes number of packets in queue (backlog) in n-th time slot
- $\mathbb{A}_n$ = nr. of packet arrivals, $\mathbb{D}_n$ = nr. of departures (during n-th slot)
- If the queue is empty $Q_n = 0$ then there are no departures

  $\Rightarrow$ Queue length at time $n+1$ can be written as

$$Q_{n+1} = Q_n + \mathbb{A}_n, \quad \text{if } Q_n = 0$$

- If $Q_n > 0$, departures and arrivals may happen

$$Q_{n+1} = Q_n + \mathbb{A}_n - \mathbb{D}_n, \quad \text{if } Q_n > 0$$

- $\mathbb{A}_n \in \{0, 1\}$, $\mathbb{D}_n \in \{0, 1\}$ and either $\mathbb{A}_n = 1$ or $\mathbb{D}_n = 1$ but not both

  $\Rightarrow$ Arrival and departure probabilities are

$$\mathrm{P}(\mathbb{A}_n = 1) = \lambda, \qquad \mathrm{P}(\mathbb{D}_n = 1) = \mu$$

Queue evolution probabilities (reminder)

- Future queue lengths depend on current length only
- Probability of queue length increasing

$$\mathrm{P}\left(Q_{n+1} = i+1 \,\middle|\, Q_n = i\right) = \mathrm{P}\left(\mathbb{A}_n = 1\right) = \lambda, \qquad \text{for all } i$$

- Queue length might decrease only if $Q_n > 0$. Probability is

$$\mathrm{P}\left(Q_{n+1} = i-1 \,\middle|\, Q_n = i\right) = \mathrm{P}\left(\mathbb{D}_n = 1\right) = \mu, \qquad \text{for all } i > 0$$

- Queue length stays the same if it neither increases nor decreases

$$\mathrm{P}\left(Q_{n+1} = i \,\middle|\, Q_n = i\right) = 1 - \lambda - \mu, \qquad \text{for all } i > 0$$
$$\mathrm{P}\left(Q_{n+1} = 0 \,\middle|\, Q_n = 0\right) = 1 - \lambda$$

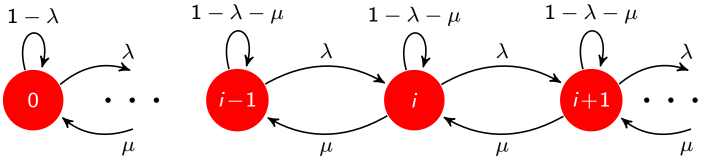
$\Rightarrow$ No departures when $Q_n = 0$ explain second equation

Queue as a Markov chain (reminder)

- MC with states $0, 1, 2, \ldots$. Identify states with queue lengths
- Transition probabilities for $i \neq 0$ are

$$P_{i,i-1} = \mu, \qquad P_{i,i} = 1 - \lambda - \mu, \qquad P_{i,i+1} = \lambda$$

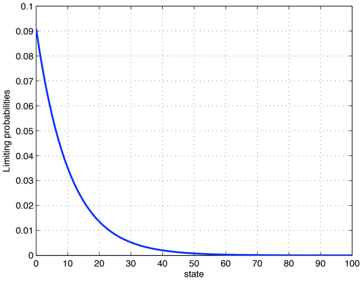
- For $i = 0$: $P_{00} = 1 - \lambda$ and $P_{01} = \lambda$



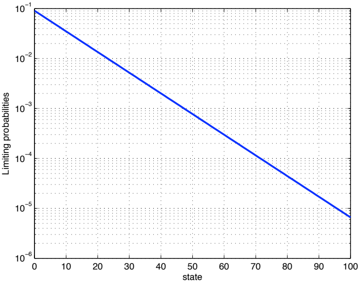
Numerical example: Limit probabilities

- Build matrix $\mathbf{P}$ truncating at maximum queue length $L = 100$
  - $\Rightarrow$ Arrival rate $\lambda = 0.3$. Departure rate $\mu = 0.33$
- Find eigenvector of $\mathbf{P}^T$ associated with eigenvalue 1
  - $\Rightarrow$ Yields limit probabilities $\boldsymbol{\pi} = \lim_{n\to\infty} \mathbf{p}(n)$ (ergodic MC)

linear scale

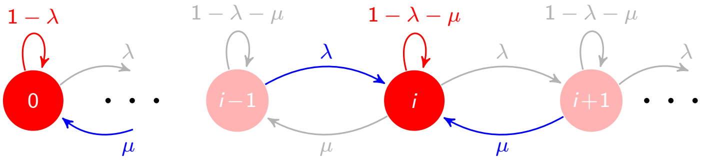


logarithmic scale



- Limit probabilities appear linear in logarithmic scale
  - $\Rightarrow$ Seemingly implying an exponential expression $\pi_i \propto \alpha^i$ $(0 < \alpha < 1)$

Limit distribution equations



- Total probability yields

$$\mathrm{P}\left(X_{n+1}=i\right)=\sum_{j=i-1}^{i+1}\mathrm{P}\left(X_{n+1}=i \,\middle|\, X_n=j\right)\mathrm{P}\left(X_n=j\right)$$

- Limit distribution equations for state 0 (empty queue)

$$\pi_0=(1-\lambda)\pi_0+\mu\pi_1$$

- For the remaining states $i \neq 0$

$$\pi_i=\lambda\pi_{i-1}+(1-\lambda-\mu)\pi_i+\mu\pi_{i+1}$$

Verification of candidate solution

- Substitute candidate solution $\pi_i = c\alpha^i$ in equation for π_0

$$c\alpha^0 = (1-\lambda)c\alpha^0 + \mu c\alpha^1 \quad \Rightarrow \quad 1 = (1-\lambda) + \mu\alpha$$

$\Rightarrow$ The above equation holds for $\alpha = \lambda/\mu$

- Q: Does $\alpha = \lambda/\mu$ verify the remaining equations?
- From the equation for generic π_i (divide by $c\alpha^{i-1}$)

$$c\alpha^i = \lambda c\alpha^{i-1} + (1-\lambda-\mu)c\alpha^i + \mu c\alpha^{i+1}$$
$$\mu\alpha^2 - (\lambda+\mu)\alpha + \lambda = 0$$

$\Rightarrow$ The above quadratic equation is satisfied by $\alpha = \lambda/\mu$

$\Rightarrow$ And $\alpha = 1$, which is irrelevant

Compute normalization constant

- Next, determine c so that probabilities sum to 1 ($\sum_{i=0}^{\infty} \pi_i = 1$)

$$\sum_{i=0}^{\infty} \pi_i = \sum_{i=0}^{\infty} c(\lambda/\mu)^i = \frac{c}{1 - \lambda/\mu} = 1$$

  $\Rightarrow$ Used geometric sum, need $\lambda/\mu < 1$ (queue stability condition)

- Solving for c and substituting in $\pi_i = c\alpha^i$ yields

$$\pi_i = (1 - \lambda/\mu) \left(\frac{\lambda}{\mu}\right)^i$$

- The ratio μ/λ is the queue's stability margin

  $\Rightarrow$ Probability of having fewer queued packets grows with μ/λ

Queue balance equations

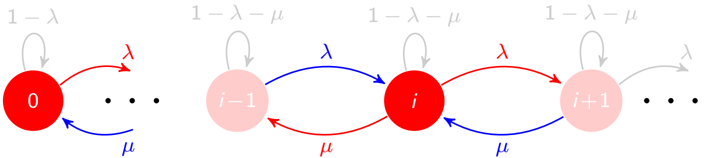
- Rearrange terms in equation for limit probabilities [cf. page 38]

$$\lambda\pi_0 = \mu\pi_1$$
$$(\lambda+\mu)\pi_i = \lambda\pi_{i-1} + \mu\pi_{i+1}$$

  - $\lambda\pi_0$ is average rate at which the queue leaves state 0
  - Likewise $(\lambda+\mu)\pi_i$ is the rate at which the queue leaves state i
  - $\mu\pi_1$ is average rate at which the queue enters state 0
  - $\lambda\pi_{i-1} + \mu\pi_{i+1}$ is rate at which the queue enters state i

- Limit equations prove validity of queue balance equations

$$\text{Rate at which leaves} = \text{Rate at which enters}$$



Concurrent arrival and departures

- Packets may arrive and depart in same time slot (concurrence)
  - $\Rightarrow$ Queue evolution equations remain the same [cf. page 34]
  - $\Rightarrow$ But queue probabilities change [cf. page 35]

- Probability of queue length increasing (for all i)

$$\mathrm{P}\left(Q_{n+1} = i+1 \,\middle|\, Q_n = i\right) = \mathrm{P}\left(\mathbb{A}_n = 1\right)\mathrm{P}\left(\mathbb{D}_n = 0\right) = \lambda(1-\mu)$$

- Queue length might decrease only if $Q_n > 0$ (for all $i > 0$)

$$\mathrm{P}\left(Q_{n+1} = i-1 \,\middle|\, Q_n = i\right) = \mathrm{P}\left(\mathbb{A}_n = 0\right)\mathrm{P}\left(\mathbb{D}_n = 1\right) = (1-\lambda)\mu$$

- Queue length stays the same if it neither increases nor decreases

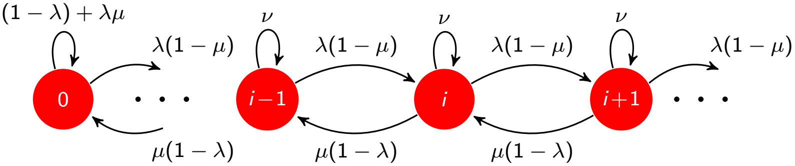
$$\mathrm{P}\left(Q_{n+1} = i \,\middle|\, Q_n = i\right) = \lambda\mu + (1-\lambda)(1-\mu) = \nu, \qquad \text{for all } i > 0$$
$$\mathrm{P}\left(Q_{n+1} = 0 \,\middle|\, Q_n = 0\right) = (1-\lambda) + \lambda\mu$$

Limit distribution from queue balance equations

- Write limit distribution equations $\Rightarrow$ Queue balance equations

  $\Rightarrow$ Rate at which leaves $=$ Rate at which enters

$$\lambda(1-\mu)\pi_0 = \mu(1-\lambda)\pi_1$$
$$\big(\lambda(1-\mu)+\mu(1-\lambda)\big)\pi_i = \lambda(1-\mu)\pi_{i-1} + \mu(1-\lambda)\pi_{i+1}$$



- Again, try an exponential solution $\pi_i = c\alpha^i$

Solving for limit distribution

- Substitute candidate solution in equation for π_0

$$\lambda(1-\mu)c = \mu(1-\lambda)c\alpha \quad \Rightarrow \quad \alpha = \frac{\lambda(1-\mu)}{\mu(1-\lambda)}$$

- Same substitution in equation for generic π_i

$$\mu(1-\lambda)c\alpha^2 + \big(\lambda(1-\mu) + \mu(1-\lambda)\big)c\alpha + \lambda(1-\mu)c = 0$$

$\Rightarrow$ As before is solved for $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$

- Find constant c to ensure $\sum_{i=0}^{\infty} c\alpha^i = 1$ (geometric series). Yields

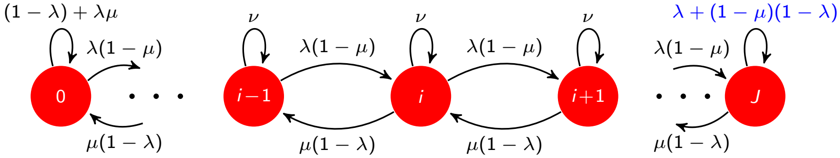
$$\pi_i = (1-\alpha)\alpha^i = \left(1 - \frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)\left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i$$

Limited queue size

- Packets dropped if queue backlog exceeds buffer size J

  $\Rightarrow$ Many packets $\rightarrow$ large delays $\rightarrow$ packets useless upon arrival

  $\Rightarrow$ Also preserve memory



- Should modify equation for state J (Rate leaves $=$ Rate enters)

$$\mu(1-\lambda)\pi_J = \lambda(1-\mu)\pi_{J-1}$$

- $\pi_i = c\alpha^i$ with $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$ also solves this equation (Yes!)

Compute limit distribution

- Limit probabilities are not the same because constant c is different
- To compute c, sum a finite geometric series

$$1 = \sum_{i=0}^{J} c\alpha^i = c\frac{1-\alpha^{J+1}}{1-\alpha} \quad \Rightarrow \quad c = \frac{1-\alpha}{1-\alpha^{J+1}}$$

- Limit probabilities for the finite queue thus are

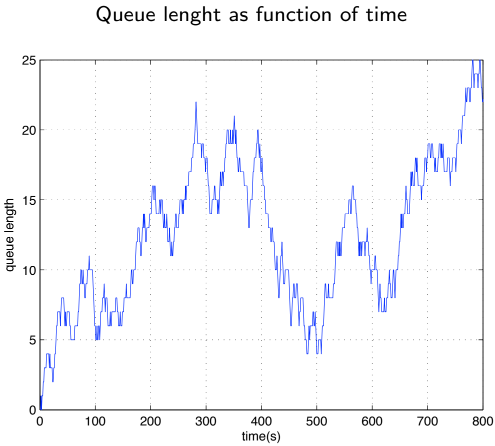
$$\pi_i = \frac{1-\alpha}{1-\alpha^{J+1}}\alpha^i \approx (1-\alpha)\alpha^i$$

$\Rightarrow$ Recall $\alpha = \lambda(1-\mu)/\mu(1-\lambda)$, and $\approx$ valid for large J

- Large J approximation yields same result as infinite length queue

Simulations: Process realization

- Arrival rate $\lambda = 0.3$. Departure rate $\mu = 0.33$. Resulting $\alpha \approx 0.87$
- Maximum queue length $J = 100$. Initial state $Q_0 = 0$ (queue empty)



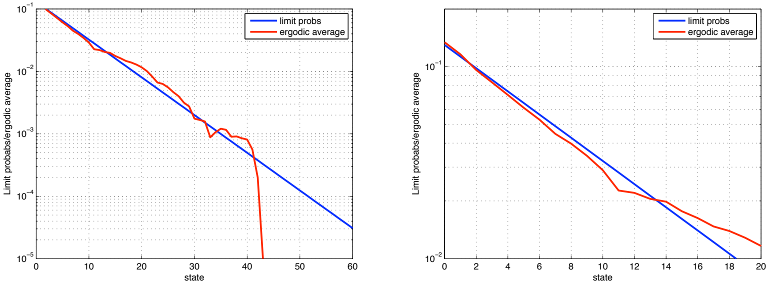
Simulations: Average occupancy and limit distribution

- Can estimate average time spent at each queue state

  $\Rightarrow$ Should coincide with the limit (stationary) distribution $\boldsymbol{\pi}$

60 states

20 states



- For $i = 60$ occupancy probability is $\pi_i \approx 10^{-5}$

  $\Rightarrow$ Explains inaccurate prediction for large i (rarely visit state i)

Buffer overflow

- Closing the loop, recall our buffer design problem
  - Arrival rate $\lambda = 0.45$ and departure rate $\mu = 0.55$
  - How big should the buffer be to have a drop rate smaller than 10^{-6}? (i.e., one packet dropped for every million packets handled)

- Q: What is the probability of buffer overflow (non-concurrent case)?
- A: Packet discarded if queue is in state J and a new packet arrives

$$\mathrm{P}\left(\text{overflow}\right) = \lambda\pi_J = \frac{1-\alpha}{1-\alpha^{J+1}}\lambda\alpha^J \approx (1-\alpha)\lambda\alpha^J$$

$\Rightarrow$ With $\lambda = 0.45$ and $\mu = 0.55$, $\alpha \approx 0.82$ $\Rightarrow$ $J \approx 57$

- A final caveat

  $\Rightarrow$ Still assuming only 1 packet arrives per time slot

  $\Rightarrow$ Lifting this assumption requires continuous-time MCs

Glossary

- Periodicty
- Aperiodic state
- Positive recurrent state
- Null recurrent state
- Ergodic state
- Limit probabilities
- Stationary distribution
- Ergodic average
- Ensemble average
- Oscillating probabilities
- Reducible Markov chain
- Ergodic component
- Non-concurrent queue
- Queue limit probabilities
- Queue stability condition
- Stability margin
- Balance equations
- Concurrency
- Limited queue size
- Buffer overflow